Forecasting and Risk (BANA 4090)

ARIMA models (Part I)

Zhaohu(Jonathan) Fan

06/29/2021

- Main topics:
    - ARIMA models
        - Time series patterns
        - Stationarity and differencing

Prerequisites

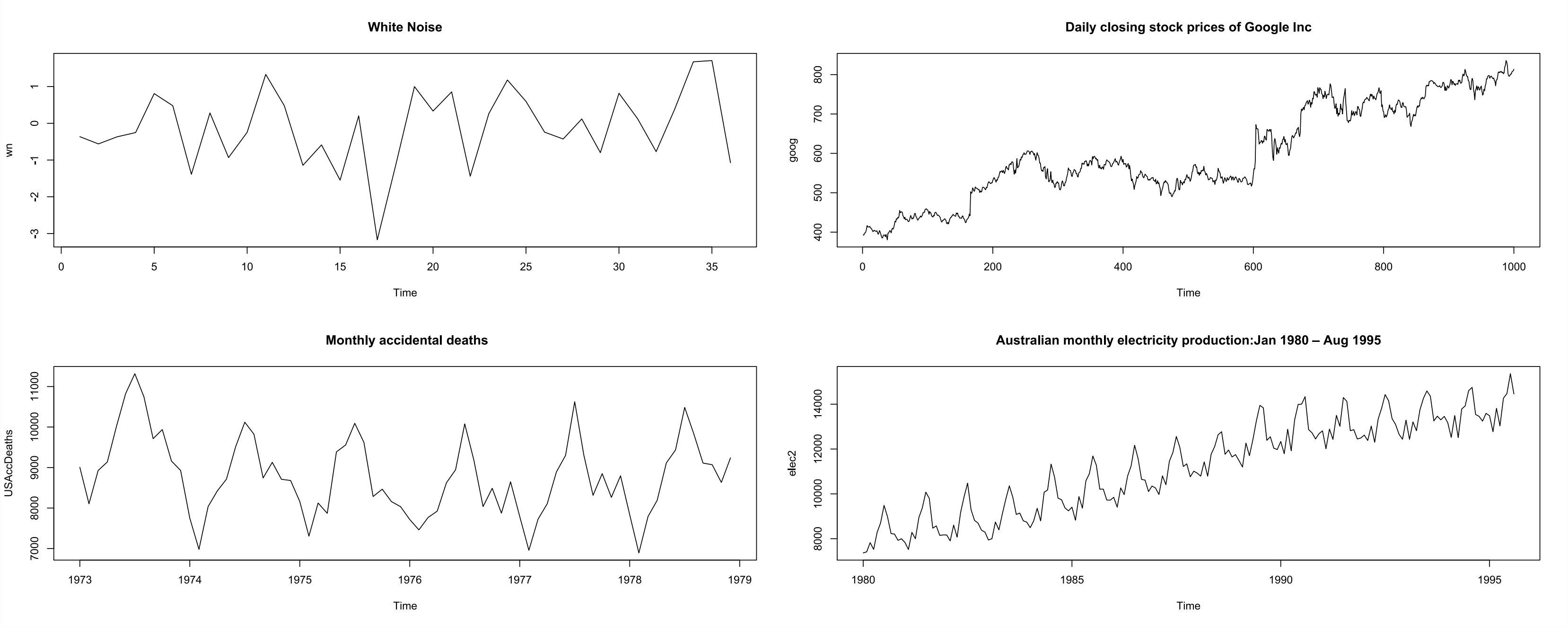
```r
# List of required (CRAN) packages
pkgs ← c(
  "ggplot2",# for drawing nicer graphics
  "fpp2",   # for using four simple forecasting models
  "forecast", #for using checkresiduals() function: a test of autocorrelation of the
)

# Install required (CRAN) packages
for (pkg in pkgs) {
  if (!(pkg %in% installed.packages()[, "Package"])) {
    install.packages(pkg)
  }
}
```

Time series patterns

Time series patterns

- Can you spot any seasonality, cyclicity and trend?
- What do you learn about the series?



Autocorrelation (ACF)

Autocorrelation: measure linear relationship between **lagged values** of a time series y.

We measure the relationship between:

- y_t and y_{t-1}
- y_t and y_{t-2}
- y_t and y_{t-3}
- etc.

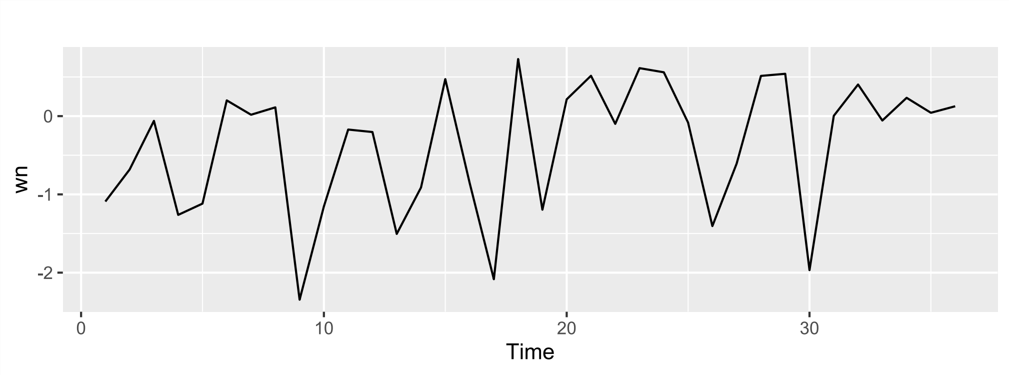
Time series patterns

We will explore time series by using the following graphics functions (in `R`):

- `ggAcf`

Example: White noise

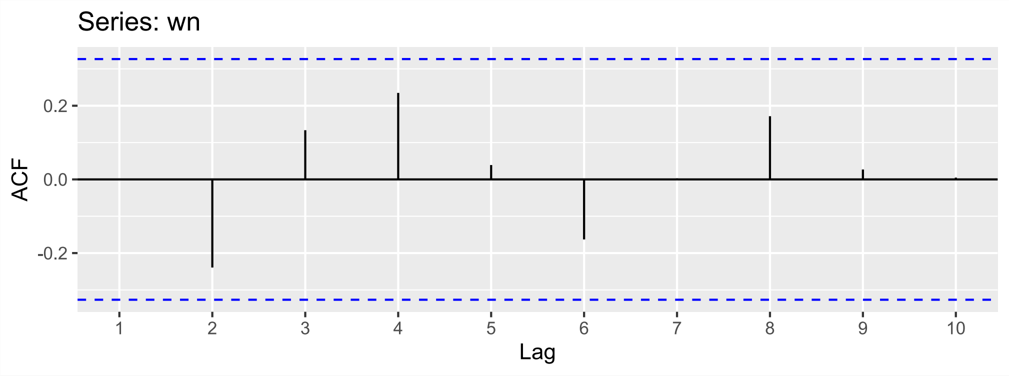
```r
wn <- ts(rnorm(36))
autoplot(wn)
```



Example: White noise

- Sample autocorrelations for white noise series.

```
set.seed(2021)
ggAcf(wn,ci=0.95, lag.max=10)
```



- We expect each autocorrelation to be close to zero.

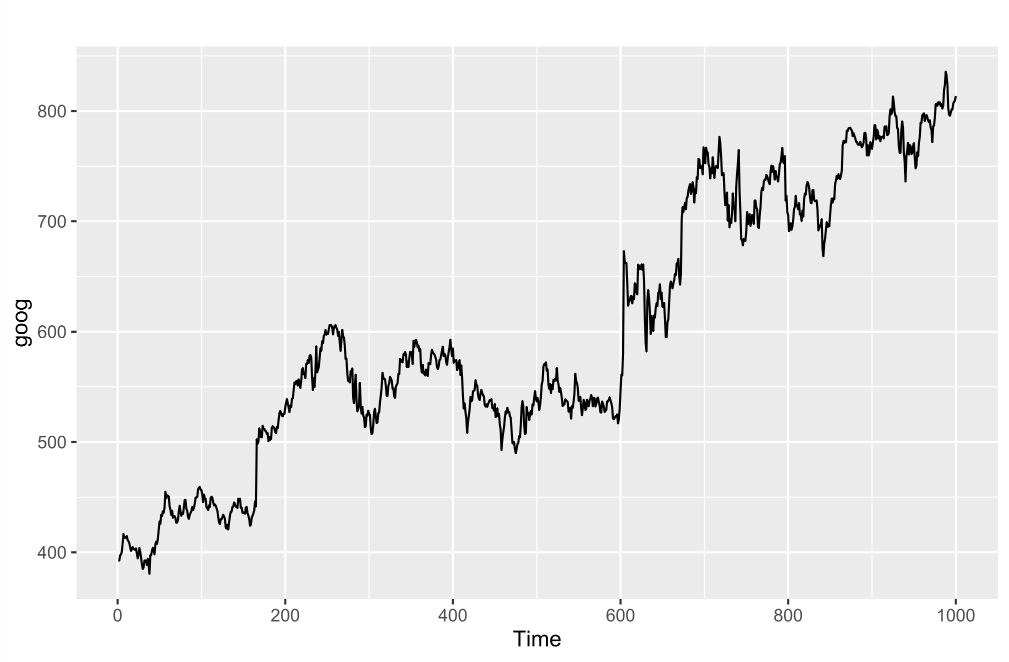
Trend and seasonality in ACF plots

- When data have a **trend**, the autocorrelations for small lags tend to be large and positive.
- When data are **seasonal**, the autocorrelations will be larger at the seasonal lags (i.e., at multiples of the seasonal frequency)
- When data are **trended and seasonal**, you see a combination of these effects.

Case 1: Google stock price

- Daily closing stock prices of Google Inc (for 1000 consecutive trading days between 25 February 2013 and 13 February 2017).
- When data have a **trend**, the autocorrelations for small lags tend to be large and positive.

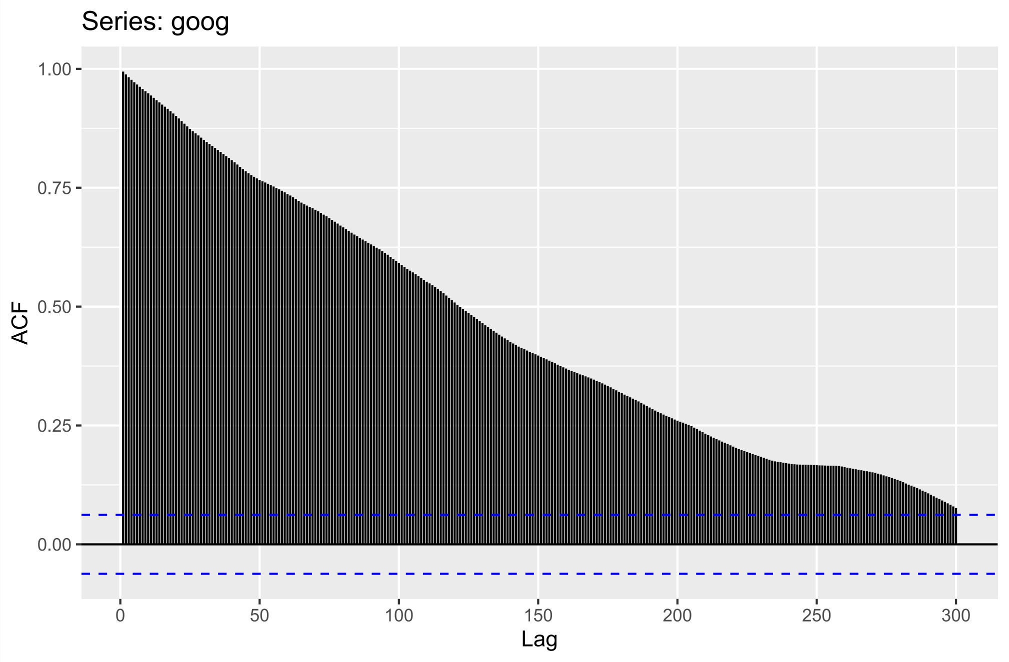
```
autoplot(goog)
```



Case 1: Google stock price (cont'd)

- When data have a **trend**, the autocorrelations for small lags tend to be large and positive.

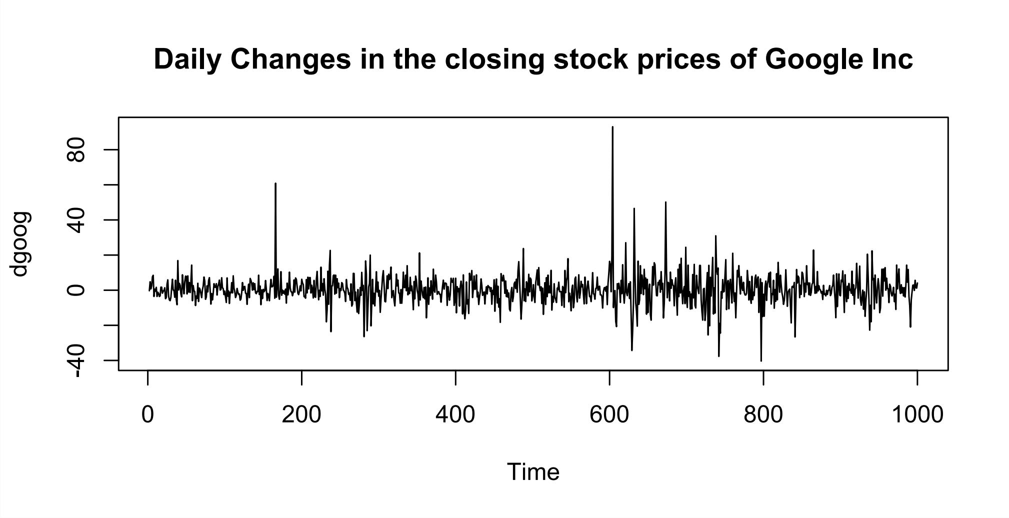
```
ggAcf(goog, lag.max=300)
```



Case 1: Google stock price (Differencing)

- Differencing helps to **stabilize the mean**.
- The differenced series is the change between each observation in the original series: $y'_t = y_t - y_{t-1}$.
- The differenced series will have only $T - 1$ values since it is not possible to calculate a difference y'_1 for the first observation.

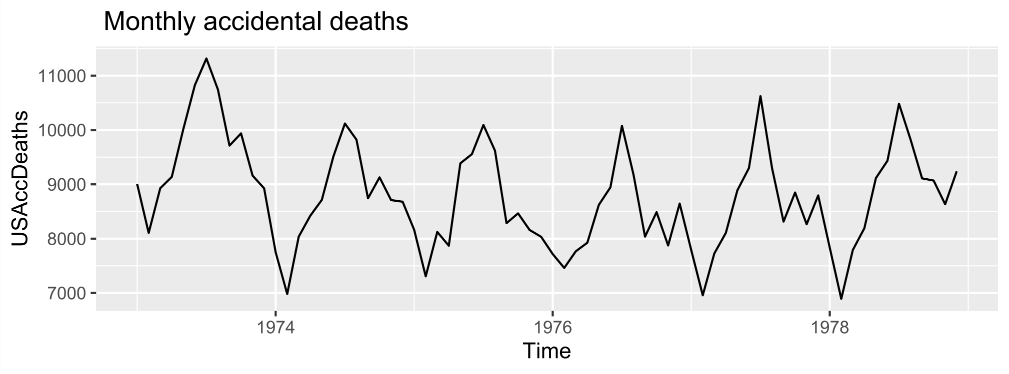
```
dgoog<-diff(goog)
plot(dgoog,main="Daily Changes in the closing stock prices of Google Inc")
```



Case 2: Accidental Deaths in the US

- A time series giving the monthly totals of accidental deaths in the USA 1973–1978.
- The values for the first six months of 1979 are 7798 7406 8363 8460 9217 9316.

```
autoplot(USAccDeaths ,main=" Monthly accidental deaths")
```



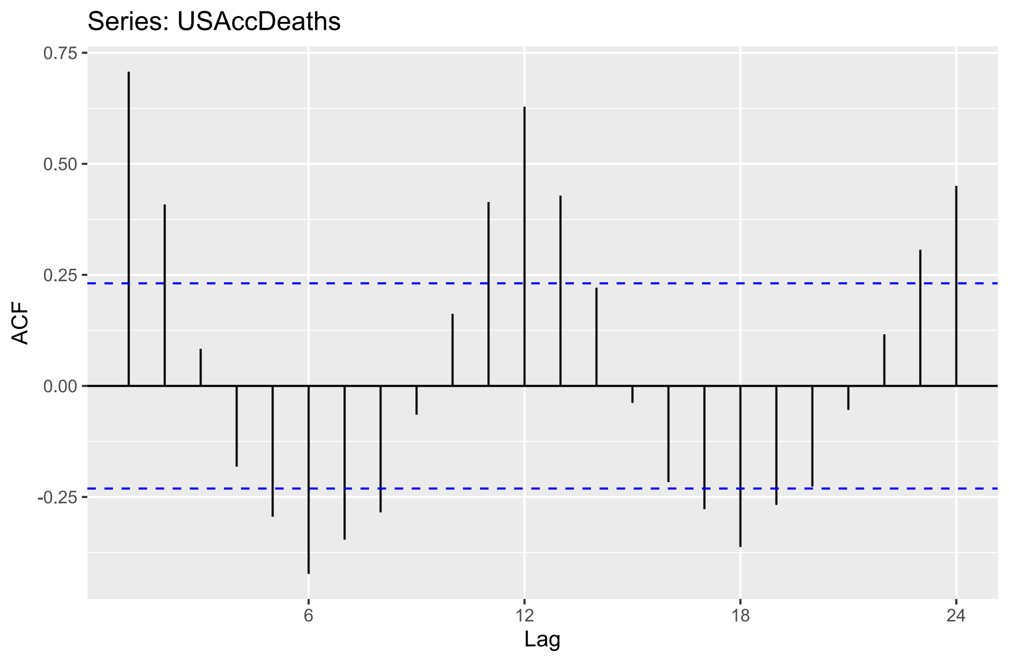

```
head(USAccDeaths)
```

```
##          Jan    Feb    Mar    Apr    May    Jun
## 1973    9007   8106   8928   9137  10017  10826
```

Case 2: Accidental Deaths in the US

- When data are **seasonal**, the autocorrelations will be larger at the seasonal lags (i.e., at multiples of the seasonal frequency).
- The ACF peaks at lags 6, 12, 18, …, indicate seasonality of length 6.

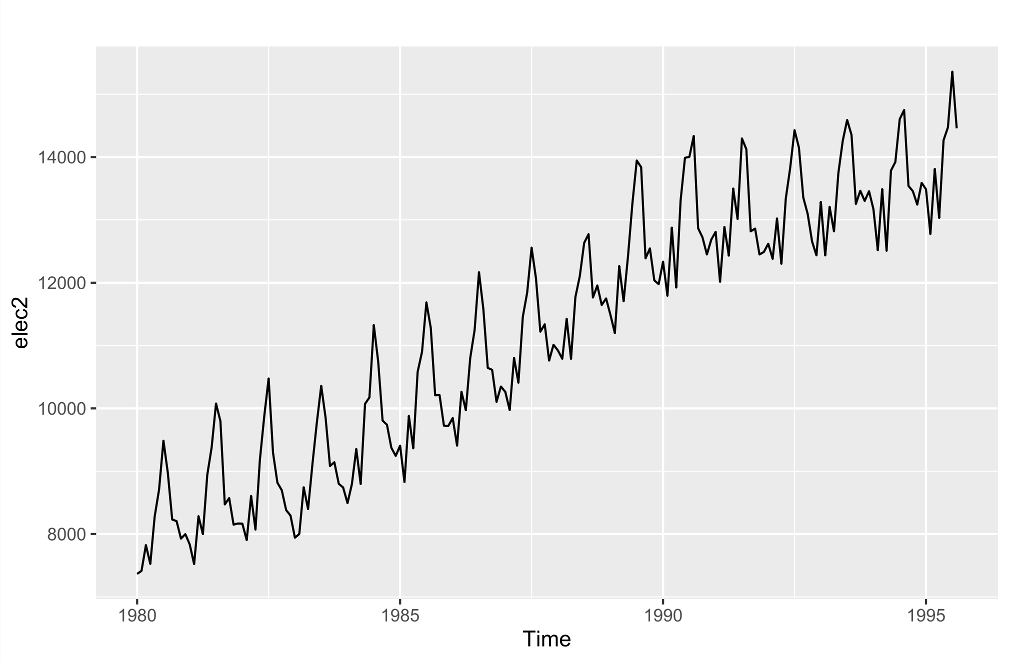
```
ggAcf(USAccDeaths)
```



Case 3: Australian monthly electricity

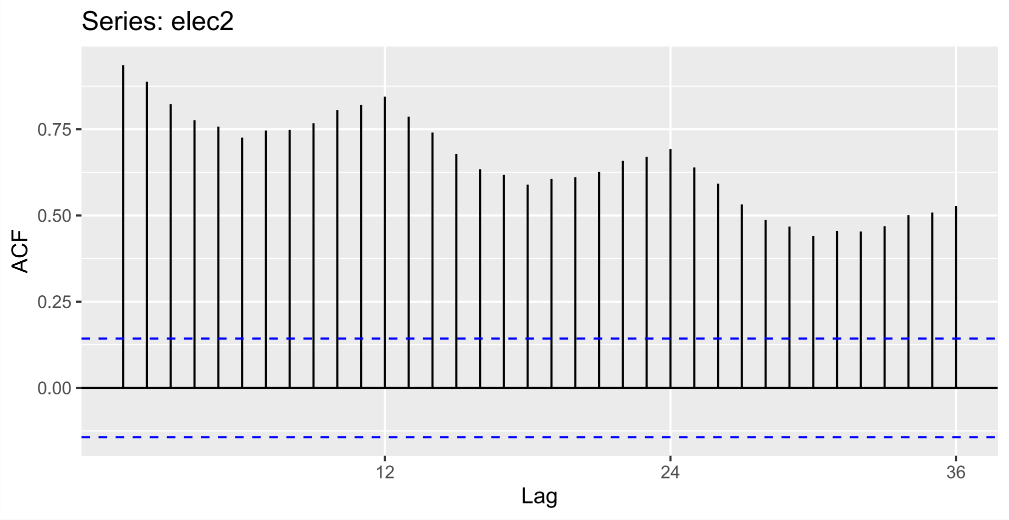
- Australian monthly electricity production: Jan 1980 – Aug 1995.
- When data are **trended and seasonal**, you can see a combination of these effects.

```
elec2 ← window(elec, start=1980)
autoplot(elec2)
```



Case 3: Australian monthly electricity

```
ggAcf(elec2, lag.max=36)
```



- Time plot shows clear **trend and seasonality**.
- The same features are reflected in the ACF.
  - The slowly decaying ACF indicates trend.
  - The ACF peaks at lags 12, 24, 36, . . ., indicate seasonality of length 12.

Summary: trend and seasonality in ACF

- When data have a **trend**, the autocorrelations for small lags tend to be large and positive.
- When data are **seasonal**, the autocorrelations will be larger at the seasonal lags (i.e., at multiples of the seasonal frequency)
- When data are **trended and seasonal**, you see a combination of these effects.